

Classification tree for the prediction of malignant disease and the prediction of non-diagnostic biopsies in patients with small renal masses

Michael Organ¹; Landan P. MacDonald²; Michael A.S. Jewett³; Henry Ajzenberg⁴; Ashraf Almatar⁵; Mohamed Abdoell⁶; Matthew R. Acker⁷; Ricardo Rendon⁸

¹Department of Urology, Memorial University, St. John's, Newfoundland, Canada; ²School of Medicine, Dalhousie University, Saint John, NB, Canada; ³Departments of Surgery (Urology) and Surgical Oncology, Princess Margaret Cancer Centre, University Health Network and University of Toronto, Toronto, ON, Canada; ⁴School of Medicine, Queen's University, Kingston, ON, Canada; ⁵Department of Urology, King Fahad Specialist Hospital-Dammam, Dammam, Saudi Arabia; ⁶Department of Diagnostic Radiology, Dalhousie University, Halifax, NS, Canada; ⁷Department of Urology, Dalhousie University, Saint John, NB, Canada; ⁸Department of Urology, Dalhousie University, Halifax, NS, Canada

Cite as: *Can Urol Assoc J* 2018 September 27; Epub ahead of print.

<http://dx.doi.org/10.5489/cuaj.5196>

Published online September 27, 2018

Abstract

Introduction: Preoperative prediction of benign vs. malignant small renal masses (SRMs) remains a challenge. This study 1) validates our previously published classification tree (CT) with an external cohort; (2) creates a new CT with the combined cohort; and 3) evaluates the R.E.N.A.L. and PADUA scoring systems for prediction of malignancy.

Methods: This study includes a total of 818 patients with renal masses; 395 underwent surgical resection and 423 underwent biopsy. A CT to predict benign disease was developed using patient and tumour characteristics from the 709 eligible participants. Our CT is based on four parameters: tumour volume, symptoms, gender, and symptomatology. CART modelling was also used to determine if R.E.N.A.L. and PADUA scoring could predict malignancy.

Results: When externally validated with the surgical cohort, the predictive accuracy of the old CT dropped. However, by combining the cohorts and creating a new CT, the predictive accuracy increased from 74% to 87% (95% confidence interval 0.84–0.89). R.E.N.A.L. and PADUA score alone were not predictive of malignancy. One limitation was the lack of available histological data from the biopsy series.

Conclusions: The validated old CT and new combined-cohort CT have a predictive value greater than currently published nomograms and single-biopsy cohorts. Overall, R.E.N.A.L. and PADUA scores were not able to predict malignancy.

Introduction

Renal cell carcinoma (RCC) is the third most common urologic malignancy.¹ Its incidence has been increasing due to increased detection of small renal masses (SRMs, defined as <4 cm).² The majority of SRMs are asymptomatic at diagnosis and have a non-aggressive behavior,^{3,4} however some (up to 6.0%) can present with metastasis.^{5,6} Curative therapy including partial and radical nephrectomy are associated with considerable morbidity.⁷⁻⁹

Small renal masses are found to be histologically benign in 20-46% of cases.¹⁰⁻¹³ This prompted investigation into methods to predict benign vs. malignant disease. We have previously developed a classification tree (CT) using patient and tumour characteristics (tumour volume, location, and symptoms) to predict benign vs. malignant disease with an overall accuracy of 89%.¹⁴ (Figure 1) Two nomograms have been published that can be used as tools to predict surgical histology.^{15,16} These are complex and have not been widely adopted into clinical practice. It has been suggested that the R.E.N.A.L. scoring, which was designed to predict complications after partial nephrectomy^{17,18}, can predict benign from malignant disease, with more complex lesions having a greater malignant potential.^{16,19,20}

The aims of this study were to externally validate our previously published CT¹⁴ using patient and tumour characteristics, and create a new CT with a combined cohort from two different institutions to predict histology using tumour size, location, symptoms and patient gender. Furthermore, the R.E.N.A.L. and PADUA scoring systems were evaluated to determine if they could predict malignancy or non-diagnostic biopsies.

Methods

Local institutional review boards approved this study which includes 818 renal masses. 423 patients were treated at the Princess Margaret Cancer Centre (PM), Toronto, Canada and underwent a renal mass biopsy. 395 consecutive patients from QE2 Health Sciences Centre (QE2) in Halifax, Nova Scotia, Canada underwent surgical treatment for renal mass. The PM cohort was obtained from a prospectively maintained database of 423 patients with SRMs undergoing treatment or surveillance. All patients undergoing percutaneous SRM biopsy between January 2000 and December 2009 were eligible for inclusion in this study. Biopsies were performed for planning management (surveillance or intervention), at the time of thermal ablation or in the post-ablation period for suspicion of recurrence.

The QE2 cohort included 395 patients who had open or laparoscopic partial or radical nephrectomy for renal masses ≤ 5 cm between July 1st, 2001 and June 30th, 2010. Patients were identified from an institutionally maintained prospective database of patients with renal masses and from physician records. All patients were >18 years of age and had a renal mass with imaging characteristics consistent with RCC. Patients with renal angiomyolipoma were excluded.

We used the combined PM+QE2 cohort to determine the validity of our previously published CT¹⁴. We then used binary recursive partitioning analysis (RPA) to create the new, combined CT model for the prediction of tumour histology (benign vs. malignant), and to

evaluate R.E.N.A.L. and PADUA scoring for the same. Potential patient prognostic factors used to develop the new CT included age, sex, and symptoms at diagnosis. Each renal mass was classified broadly as either benign or malignant. Pre-treatment images were reviewed for each renal mass. Potential radiographic predictive factors included tumour volume (three dimensions: $V = \frac{4}{3}\pi \frac{xyz}{8}$; two dimensions $V = \frac{4}{3}\pi \frac{xy(x+y)}{8}$; and one dimension: $V = \frac{4}{3}\pi \frac{x^3}{8}$, tumour location (central or peripheral), degree of endophytic component (1-100%), and tumour axis location. Tumour location was defined as central if the tumour was in direct contact with or invading the collecting system and/or renal sinus. All other renal masses were defined as peripheral. Degree of endophytic component was recorded as the percentage of the tumour that was within the normal contour of the kidney. Tumour axis location was designated according to three renal axes: 1) upper pole, interpolar, or lower pole, 2) medial or lateral and 3) anterior or posterior. Binary RPA implemented via classification and regression tree (CART) methodology is a semi-parametric modeling algorithm that generates simply understood binary decision tree models that stratify cases into various risk categories according to several prognostic factors. From the perspective of knowledge translation, a CT is appealing as a clinical decision tool as it is simply represented and implemented as a set of binary decision rules. The CT models predict malignant or benign outcome as a function of a set of explanatory variables. The CT models were developed automatically using the *rpart* package in the R language for statistical computing.²¹ Sensitivity, specificity, accuracy, and positive and negative predictive values were computed for the CT models.

For the new CT, a random forest was utilized. Random forests is an ensemble method that constructs a collection of CTs with strategically injected variation by combining the concepts of bagging and the random selection of features. Random forests are effective at generating internal unbiased estimates of the generalization error. While random forests are difficult to interpret because they combine many models together, they typically outperform any single model generated by CART analysis. The random forest model was developed using the *randomForest* package in the R language for statistical computing.²² Sensitivity, specificity, accuracy, and positive and negative predictive values were obtained from the confusion matrix generated by the random forest model.

Results

The total combined cohort sample was 818 patients. Renal biopsy was initially performed in 423 PM cohort patients, 278 (66%) of which were male. 288 (68%) masses were detected incidentally. The median tumour volume was 8.7 cm³, while the median endophytic component was 50%. Only 357 (84% of the total 423 PM cohort masses) who underwent a diagnostic biopsy were included in the cohort and 281 (79%) of these revealed malignancy, the majority of which were clear cell RCC. Forty-three participants with unknown symptomatology from both PM and QE2 cohorts were excluded from the combined cohort further reducing the participants to a

final number of 709. (See **Table 1** for further demographic data). R.E.N.A.L. and PADUA scores were also calculated for all the masses in the biopsy series.

As previously reported, the old CT demonstrated an accuracy of 89%.¹⁴ When externally validated with the biopsy cohort (PM), the accuracy decreased to 74% (95% CI; 0.69-0.78). The QE2 and PM patients were then combined and the new CT was developed, which demonstrated an accuracy of 87% (95% CI; 0.84-0.89). As this new CT was not externally validated with a separate cohort, a random forest technique was used to create a more robust CT. (Figure 2) Using the CART methodology, the R.E.N.A.L. nephrometry score alone was not able to predict whether the SRM was benign or malignant. When looking individually at the components of the R.E.N.A.L. score, “nearness to the collecting system” was associated with malignancy, with those within 7mm of the collecting system (N scores 2 and 3) having a higher rate of malignancy (30%) than those further away (15%). Alone, size of the tumour was not predictive of malignancy, but when used with “nearness to the collecting system” in the form of a CT, larger masses (R scores 2 and 3) were more likely to be malignant. Neither the overall PADUA score or the individual components were predictive of malignancy. (Figure 3)

Discussion

Currently SRMs are being over-treated as not all are malignant. Pre-operative prediction of malignancy in SRMs is a significant challenge as there is currently no validated, non-invasive way to predict benign vs. malignant renal disease.

Few groups have developed predictive tools to determine the risk of benign vs. malignant disease for SRMs.^{15,16} Lane *et al.* used age, gender, radiological size at diagnosis, symptoms at presentation, and smoking history to develop a nomogram.¹⁵ Their nomogram predicted benign SRM with an AUC of 64.4%. Kutikov *et al.* developed nomograms using the R.E.N.A.L. nephrometry scoring system with an AUC of 76% in predicting malignancy and 74% in predicting low vs. high grade.¹⁶ Although both of the aforementioned studies are nomograms predicting benign masses, when comparing accuracy our classification tree performs better. The R.E.N.A.L. nomogram predicting grade has been externally validated²³ and has been shown to be unable to predict malignancy.^{24,25} This group demonstrated the only significant individual variables of the R.E.N.A.L. score were tumour size and proximity to the hilum. Despite this finding, all aspects of the R.E.N.A.L. score were included in the nomogram, making it less predictive and complex and difficult to use clinically. Neither of these two nomograms have been adopted clinically to predict benign vs. malignant disease. Mullins *et al.* used R.E.N.A.L. nephrometry scoring to predict malignancy by subdividing the score in low (4-6), intermediate (7-9) and high risk (10-12), showing that in intermediate and highly complex masses, there was an increased risk of malignancy.²⁰

The old CT had been previously created and reported using patient and tumour characteristics to predict benign vs. malignant disease to help clinicians better treat SRMs¹⁴. The current study validates this using an external biopsy cohort. The decrease in overall predictive accuracy is an expected finding when the old CT was externally validated with a different mix of

patients. The new CT created using both cohorts with a random forest algorithm had an accuracy of 87%. It is likely that each group of renal masses are quite different as the QE2 is a surgical group, while the PM group comes from a biopsy cohort. (Table 2).

We have demonstrated that this tool is better than the currently published nomograms and non-invasive tools that try and predict benign vs. malignant disease. The advantage of a CT is its ease of use as it mimics the clinicians' thought process and is easily memorized, making it more likely to be used clinically. Additionally, the diagnostic accuracy of these classification trees is as high as that reported in most renal mass biopsy series.

When comparing this tool to a "biopsy all" approach we feel it may provide a personalized tactic to each individual patient. Large centers have high accuracy rates after biopsy (upwards of 80%), but it is likely that lower volume centers do not share these success rates. Also, a CT does not incur any morbidity to the patient or cost to the system.

Our new CT implies that small, symptomatic tumours are more likely to be benign. DeRoche *et al.* found that symptomatology was not associated with SRM malignancy²⁶. However the consensus shows the opposite²⁷. It is possible this counter-intuitive result is explained by statistical error. The small number of participants included in this category presented with unrelated symptoms and were incidentally found to have an SRM $<2.7 \text{ cm}^3$ ($\approx 1.7 \text{ cm}$ diameter, assuming a sphere) and were included in this category.

We evaluated the relationship of R.E.N.A.L. score and malignancy in the PM cohort. The biopsy cohort did not demonstrate a relationship with overall R.E.N.A.L. score and risk of malignancy. There was however the novel finding that nearness to the collecting system was positively correlated with malignancy.

We did not find a relationship with the overall PADUA score and malignancy. Each component was also independently assessed for a relationship with the prediction of benign or malignant masses and no association was found. These data are not surprising as R.E.N.A.L. and PADUA were created to predict complications after partial nephrectomy and not malignancy. Other studies have shown that the R.E.N.A.L. nomogram is not effective at predicting malignant potential in SRMs^{24,25}.

There are limitations to this study. The technique of CART modeling can leave out predictive factors. In each level of the CT the statistical model chooses the greatest predictor, potentially leaving out useful predictors of malignancy. The evaluation of radiographic films to determine size and location of the masses is subjective and it has been demonstrated that tumour characteristics, including size, do not have a perfect correlation, but this represents actual clinical practice. The two cohorts are different, one being a surgical series, while the other was a biopsy series which may lead to lower accuracy rates when validated. The QE2 cohort included tumours $\leq 5 \text{ cm}$, i.e. tumours were included that do not meet the criteria of SRM for the creation of the new CT. The majority of tumours were $\leq 4 \text{ cm}$, however there is still additional statistical error because of this.

Our new CT is not validated. The random forest technique helped the CT become more robust, but it is no substitute for external validation. We recommend that our new CT be externally validated, especially considering the counter-intuitive result of symptomatic SRMs with volume $<2.7\text{cm}^3$ being more likely to be benign.

Conclusions

In patients with SRMs, benign vs. malignant disease can be predicted using our new CT based on patient and tumour characteristics, with an accuracy that is greater than currently available nomograms and most biopsy series. R.E.N.A.L. and PADUA scores were not associated with the prediction of malignancy.

We propose a practical approach to evaluate SRMs using a prognostic tool such as our CT to determine which patients should receive renal mass biopsy, undergo treatment, or active surveillance.

References

1. Siegel RL, Miller KD and Jemal A: Cancer Statistics, 2017. *CA. Cancer J. Clin.* 2017; **67**: 7–30.
2. Kane CJ, Mallin K, Ritchey J, et al: Renal cell cancer stage migration: analysis of the National Cancer Data Base. *Cancer* 2008; **113**: 78–83.
3. Hollingsworth JM, Miller DC, Daignault S, et al: Rising incidence of small renal masses: a need to reassess treatment effect. *J. Natl. Cancer Inst.* 2006; **98**: 1331–1334.
4. Campbell SC, Novick AC, Belldegrun A, et al: Guideline for management of the clinical T1 renal mass. *J. Urol.* 2009; **182**: 1271–1279.
5. Lee H, Lee JK, Kim K, et al: Risk of metastasis for T1a renal cell carcinoma. *World J. Urol.* 2016; **34**: 553–559.
6. Daugherty M, Sedaghatpour D, Shapiro O, et al: The metastatic potential of renal tumors: Influence of histologic subtypes on definition of small renal masses, risk stratification, and future active surveillance protocols. *Urol. Oncol.* 2017; **35**: 153.e15–153.e20.
7. Huang WC, Levey AS, Serio AM, et al: Chronic kidney disease after nephrectomy in patients with renal cortical tumours: a retrospective cohort study. *Lancet Oncol.* 2006; **7**: 735–740.
8. Thompson RH, Boorjian SA, Lohse CM, et al: Radical nephrectomy for pT1a renal masses may be associated with decreased overall survival compared with partial nephrectomy. *J. Urol.* 2008; **179**: 468–471; discussion 472–473.
9. Berg WT, Tomaszewski JJ, Yang H, et al: Complications of Renal Surgery. *Urol. Clin. North Am.* 2017; **44**: 275–288.
10. Leveridge MJ, Finelli A, Kachura JR, et al: Outcomes of Small Renal Mass Needle Core Biopsy, Nondiagnostic Percutaneous Biopsy, and the Role of Repeat Biopsy. *Eur. Urol.* 2011.
11. Frank I, Blute ML, Cheville JC, et al: Solid renal tumors: an analysis of pathological features related to tumor size. *J. Urol.* 2003; **170**: 2217–2220.
12. Corcoran AT, Russo P, Lowrance WT, et al: A review of contemporary data on surgically resected renal masses--benign or malignant? *Urology* 2013; **81**: 707–713.
13. Chopra S, Liu J, Alemozaffar M, et al: Improving needle biopsy accuracy in small renal mass using tumor-specific DNA methylation markers. *Oncotarget* 2017; **8**: 5439–5448.
14. Rendon RA, Mason RJ, Kirkland S, et al: A classification tree for the prediction of benign versus malignant disease in patients with small renal masses. *Can. J. Urol.* 2014; **21**: 7379–7384.
15. Lane BR, Babineau D, Kattan MW, et al: A preoperative prognostic nomogram for solid enhancing renal tumors 7 cm or less amenable to partial nephrectomy. *J. Urol.* 2007; **178**: 429–434.
16. Kutikov A, Smaldone MC, Egleston BL, et al: Anatomic features of enhancing renal masses predict malignant and high-grade pathology: a preoperative nomogram using the RENAL Nephrometry score. *Eur. Urol.* 2011; **60**: 241–248.
17. Ficarra V, Novara G, Secco S, et al: Preoperative aspects and dimensions used for an anatomical (PADUA) classification of renal tumours in patients who are candidates for nephron-sparing surgery. *Eur. Urol.* 2009; **56**: 786–793.
18. Kutikov A and Uzzo RG: The R.E.N.A.L. nephrometry score: a comprehensive

- standardized system for quantitating renal tumor size, location and depth. *J. Urol.* 2009; **182**: 844–853.
19. Satasivam P, Sengupta S, Rajarubendra N, et al: Renal lesions with low R.E.N.A.L nephrometry score are associated with more indolent renal cell carcinomas (RCCs) or benign histology: findings in an Australian cohort. *BJU Int.* 2012; **109 Suppl**: 44–47.
 20. Mullins JK, Kaouk JH, Bhayani S, et al: Tumor complexity predicts malignant disease for small renal masses. *J. Urol.* 2012; **188**: 2072–2076.
 21. Abdoell M, LeBlanc M, Stephens D, et al: Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. *Stat. Med.* 2002; **21**: 3395–3409.
 22. Liaw A and Wiener M: Classification and Regression by randomForest. *R News* 2002; **2**: 18–22. Available at: <http://cran.r-project.org/doc/Rnews/>.
 23. Wang H-K, Zhu Y, Yao X-D, et al: External validation of a nomogram using RENAL nephrometry score to predict high grade renal cell carcinoma. *J. Urol.* 2012; **187**: 1555–1560.
 24. Koo KC, Yoo H, Shin TY, et al: External validation of the RENAL nephrometry score nomogram for predicting high-grade renal cell carcinoma in solid, enhancing, and small renal masses. *World J. Urol.* 2014; **32**: 249–255.
 25. Antonelli A, Furlan M, Sandri M, et al: The R.E.N.A.L. nephrometric nomogram cannot accurately predict malignancy or aggressiveness of small renal masses amenable to partial nephrectomy. *Clin. Genitourin. Cancer* 2014; **12**: 366–372.
 26. DeRoche T, Walker E, Magi-Galluzzi C, et al: Pathologic characteristics of solitary small renal masses: can they be predicted by preoperative clinical parameters? *Am. J. Clin. Pathol.* 2008; **130**: 560–564.
 27. Crestani A, Rossanese M, Calandriello M, et al: Introduction to small renal tumours and prognostic indicators. *Int. J. Surg.* 2016; **36**: 495–503.

Figures and Tables

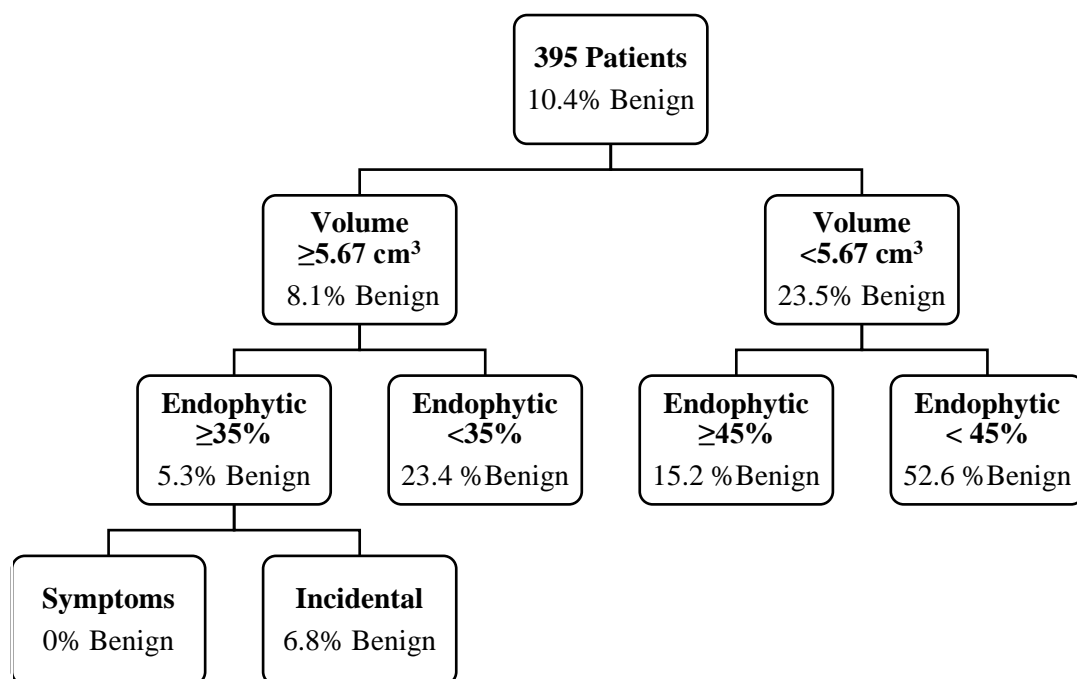
Fig. 1. Original Rendon et al classification tree predicting benign vs. malignant disease.

Fig. 2. The new, combined Organ et al classification tree using QE2 + PM patients. For each branch the absolute number of benign (B) and malignant (M) participants is provided, along with the percentage of benign participants.

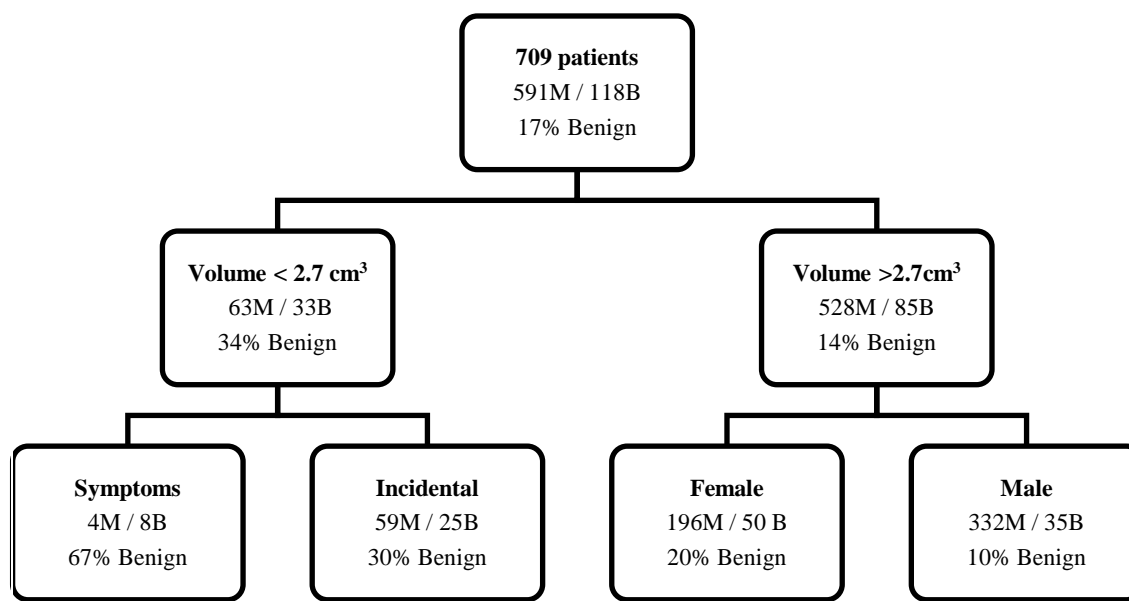


Fig. 3. Classification tree showing predictive value of the N-component of R.E.N.A.L. (nearness to the collecting system) in predicting benign vs. malignant disease. For each branch the absolute number of benign (B) and malignant (M) participants is provided along with the percentage of benign participants.

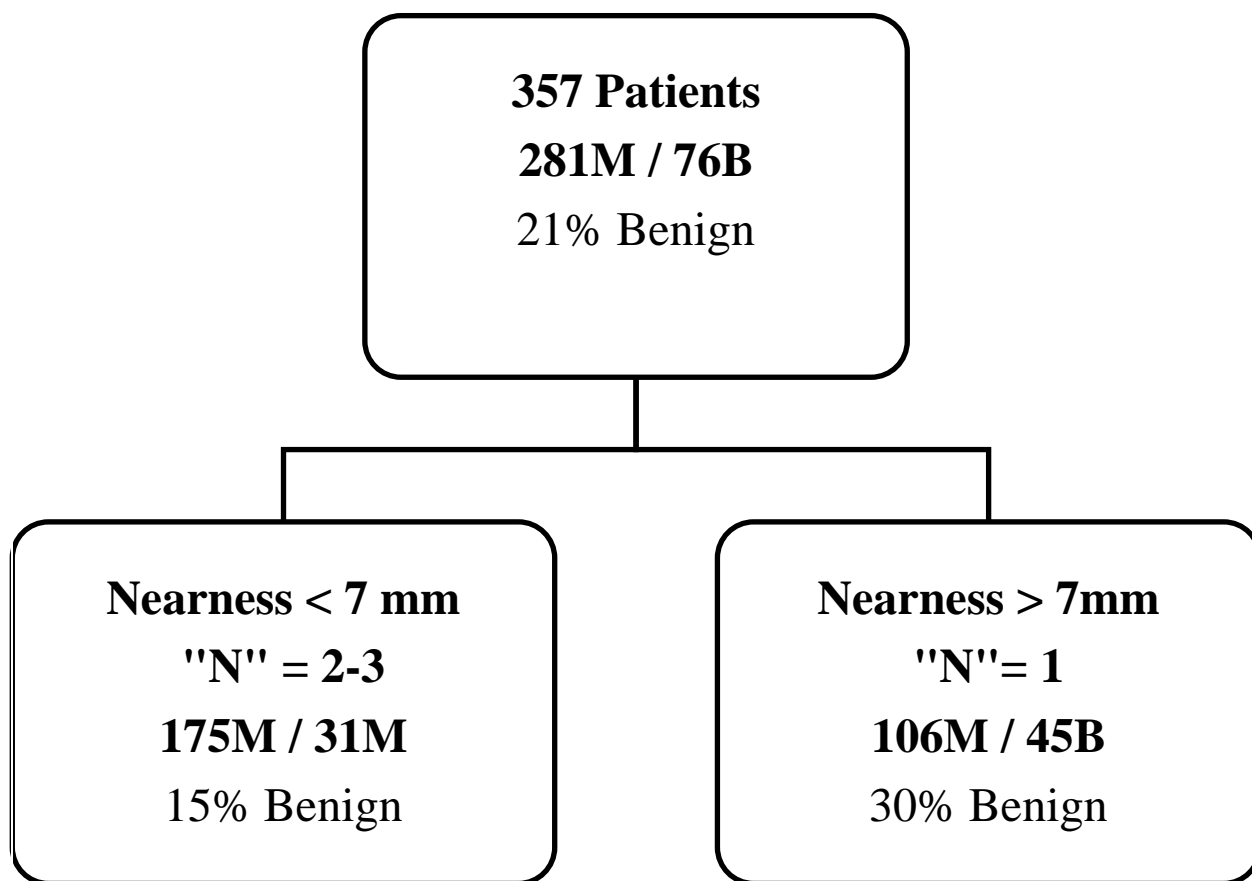


Table 1. Patient, tumour characteristics and histopathological findings		
	All PM patients	Diagnostic PM + QE2 patients
Patients	423	752
Diagnostic biopsies	357	
Non-diagnostic, n (%)	66 (16)	0
Male, n (%)	278 (66)	447 (59)
Age at surgery or biopsy, median (range)	65 (32-91)	63 (24-91)
Incidental, n (%)	288 (68)	563 (75)
Symptomatic, n (%)	84 (20)	146 (19)
Symptoms not available, n (%)	51 (12)	43 (6)
Tumour characteristics		
Imaging tumour volume in cm ³ , median (range)	8.7 (0–2252)	13 (0.8–2252)
Endophytic %, median (range)	50 (5–100)	50 (1–100)
Tumour histology		
Malignant, n (%)	281 (79)	631 (84)
Clear cell RCC, n (%)	167 (47)	431 (57)
Papillary RCC, n (%)	45 (13)	113 (15)
Chromophobe RCC, n (%)	14 (4)	31 (4)
Malignant other (%)	55 (15)	56 (7)
Benign, n (%)	76 (21)	121 (16)
Oncocytoma, n (%)	47 (13)	67 (9)
Angiomyolipoma, n (%)	16 (5)	27 (4)
Benign cystic, n (%)	1 (0)	7 (1)
Metanephric adenoma, n (%)	1 (0)	6 (1)
Leiomyoma, n (%)	0	1 (0)
Cystic nephroma, n (%)	0	1(0)
Benign other, n (%)	11 (3)	12 (2)
TMN stage		
T1a	355	527
T1b	58	141
T2a	17	13
T2b	13	13
T3a	0	13

PM: Princess Margaret; QE: Queen Elizabeth; RCC: renal cell carcinoma.

Table 2. Descriptive statistics of original and combined classification tree before and after validation		
Cohort		External validation
QE2		
Accuracy (95% CI)	0.89 (0.85, 0.92)	74 (0.69, 0.79)
Sensitivity	0.97	0.93
Specificity	0.22	0.05
Positive predictive value	0.91	0.78
Negative predictive value	0.53	0.17
Combined QE2+PM		
Accuracy (95% CI)	0.87 (0.84, 0.89)	
Sensitivity	0.94	
Specificity	0.46	
Positive predictive value	0.90	
Negative predictive value	0.61	

CI: confidence interval; PM: Princess Margaret; QE: Queen Elizabeth.